PREFACE

Our primary motivation in writing this book is to share our working experience to bridge the gap between the knowledge of industry gurus and newcomers to the spoken language processing community. Many powerful techniques hide in conference proceedings and academic papers for years before becoming widely recognized by the research community or the industry. We spent many years pursuing spoken language technology research at Carnegie Mellon University before we started spoken language R&D at Microsoft. We fully understand that it is by no means a small undertaking to transfer a state of the art spoken language research system into a commercially viable product that can truly help people improve their productivity. Our experience in both industry and academia is reflected in the context of this book, which presents a contemporary and comprehensive description of both theoretic and practical issues in spoken language processing. This book is intended for people of diverse academic and practical backgrounds. Speech scientists, computer scientists, linguists, engineers, physicists and psychologists all have a unique perspective to spoken language processing. This book will be useful to all of these special interest groups.

Spoken language processing is a diverse subject that relies on knowledge of many levels, including acoustics, phonology, phonetics, linguistics, semantics, pragmatics, and discourse. The diverse nature of spoken language processing requires knowledge in computer science, electrical engineering, mathematics, syntax, and psychology. There are a number of excellent books on the sub-fields of spoken language processing, including speech recognition, text to speech conversion, and spoken language understanding, but there is no single book that covers both theoretical and practical aspects of these sub-fields and spoken language interface design. We devote many chapters systematically introducing fundamental theories needed to understand how speech recognition, text to speech synthesis, and spoken

language understanding work. Even more important is the fact that the book highlights what works well in practice, which is invaluable if you want to build a practical speech recognizer, a practical text to speech synthesizer, or a practical spoken language system. Using numerous real examples in developing Microsoft's spoken language systems, we concentrate on showing how the fundamental theories can be applied to solve real problems in spoken language processing.

We would like to thank many people who helped us during our spoken language processing R&D careers. We are particularly indebted to Professor Raj Reddy at the School of Computer Science, Carnegie Mellon University. Under his leadership, Carnegie Mellon University has become a center of research excellence on spoken language processing. Today's computer industry and academia benefited tremendously from his leadership and contributions.

Special thanks are due to Microsoft for its encouragement of spoken language R&D. The management team at Microsoft has been extremely generous to our. We are particularly grateful to Bill Gates, Nathan Myhrvold, Rick Rashid, Dan Ling, and Jack Breese for the great environment they created for us at Microsoft Research.

Scott Meredith helped us writing a number of chapters in this book and deserves to be a co-author. His insight and experience to text to speech synthesis enriched this book a great deal. We also owe gratitude to many colleagues we worked with in the speech technology group of Microsoft Research. In alphabetic order, Bruno Alabiso, Fil Alleva, Ciprian Chelba, James Droppo, Doug Duchene, Li Deng, Joshua Goodman, Mei-Yuh Hwang, Derek Jacoby, Y.C. Ju, Li Jiang, Ricky Loynd, Milind Mahajan, Peter Mau, Salman Mughal, Mike Plumpe, Scott Quinn, Mike Rozak, Gina Venolia, Kuansan Wang, and Ye-Yi Wang, not only developed many algorithms and systems described in this book, but also helped to shape our thoughts from the very beginning.

In addition to those people, we want to thank Les Atlas, Alan Black, Jeff Bilmes, David Caulton, Eric Chang, Phil Chou, Dinei Florencio, Allen Gersho, Francisco Gimenez-Galanes, Hynek Hermansky, Kai-Fu Lee, Henrique Malvar, Mari Ostendorf, Joseph Pentheroudakis, Tandy Trower, Wayne Ward, and Charles Wayne. They provided us with many wonderful comments to refine this book. Tim Moore and Russ Hall at Prentice Hall helped us finish this book in a finite amount of time.

Finally, writing this book was a marathon that could not have been finished without the support of our spouses Yingzhi, Donna, and Phen, during the many evenings and weekends we spent on this project.

Redmond, WA   Xuedong Huang
October 2000   Alejandro Acero
                                                     Hsiao-Wuen Hon